european corporate governance institute

ESG Rating Disagreement and Stock Returns

Finance Working Paper N° 651/2020
August 2021

Rajna Gibson Brandon
University of Geneva, GFRI, and ECGI

Philipp Krueger
University of Geneva, GFRI, SFI and ECGI

Peter S. Schmidt
University of Geneva and GFRI

© Rajna Gibson Brandon, Philipp Krueger and Peter S. Schmidt 2021. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

This paper can be downloaded without charge from:
http://ssrn.com/abstract_id=3433728

www.ecgi.global/content/working-papers

european corporate governance institute

ECGI Working Paper Series in Finance

ESG Rating Disagreement and Stock Returns

Working Paper N° 651/2020

August 2021

Rajna Gibson Brandon
Philipp Krueger
Peter S. Schmidt

We thank the editor (William Goetzmann), the co-editor (Luis Garcia-Feijoo), and two anonymous referees for their insightful comments. We also thank Stephen Brown and Oghuzan Karakas for valuable comments. The paper also benefited from comments received during the PRI conference in Paris in September 2019, the ECGI and University of Cambridge Conference "Engage or Divest? Strategies for responsible Investing" in Cambridge in October 2019, and a seminar audience at the University of Geneva. Supported by the Swiss National Science Foundation (SNSF) within the framework of the National Research Program "Sustainable Economy: resourcefriendly, future-oriented, innovative" (NRP 73). Nadine Riand provided valuable research assistance in preparing the data used in this paper. The views expressed in this paper are solely those of the authors. We also remain responsible for any errors and omissions.

© Rajna Gibson Brandon, Philipp Krueger and Peter S. Schmidt 2021. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Abstract

Using ESG ratings from seven different data providers for a sample of S&P 500 firms between 2010 and 2017, we study the relation between ESG rating disagreement and stock returns. We find that stock returns are positively related to ESG rating disagreement, suggesting a risk premium for firms with higher ESG rating disagreement. The relation is primarily driven by disagreement about the environmental dimension. We discuss the practical implications of our findings for firms' equity cost of capital as well as for investment managers and asset owners who use ESG investment strategies.

Keywords: ESG ratings, disagreement, non-financial information, stock returns, equity cost of capital, sustainable finance

JEL Classifications: G12, G24, Q01

Rajna Gibson Brandon
Professor of Finance
University of Geneva, The Geneva Finance Research Institute
40 Boulevard du Pont d'Arve
1211 Geneva 4, Switzerland
phone: +41 223 798 983
e-mail: rajna.gibson@unige.ch

Philipp Krueger
Associate Professor of Responsible Finance
University of Geneva, Geneva School of Economics and Management
40, Bd du Pont-d'Arve
1211 Geneva 4, Switzerland
phone: +41 223 798 569
e-mail: philipp.krueger@unige.ch

Peter S. Schmidt*
Researcher
University of Zurich, Department of Banking and Finance
Schönberggasse 1
Zürich, 8001, Switzerland
e-mail: peter.schmidt@unige.ch

*Corresponding Author

s:fi
RESEARCH

Swiss Finance Institute Research Paper Series N°19-67

ESG Rating Disagreement and Stock Returns

Rajna Gibson
University of Geneva

Philipp Krueger
University of Geneva and Swiss Finance Institute

Peter Steffen Schmidt
University of Zurich and University of Geneva

ESG rating disagreement and stock returns

Rajna Gibson, Philipp Krueger, and Peter S. Schmidt*

Financial Analysts Journal, *Forthcoming*

First version: February 2019

This version: July 2021

Using ESG ratings from seven different data providers for a sample of S&P 500 firms between 2010 and 2017, we study the relation between ESG rating disagreement and stock returns. We find that stock returns are positively related to ESG rating disagreement, suggesting a risk premium for firms with higher ESG rating disagreement. The relation is primarily driven by disagreement about the environmental dimension. We discuss the practical implications of our findings for firms' equity cost of capital as well as for investment managers and asset owners who use ESG investment strategies.

Keywords: ESG ratings, disagreement, non-financial information, stock returns, equity cost of capital, sustainable finance

* University of Geneva, Swiss Finance Institute, European Corporate Governance Institute, Geneva School of Economics and Management (GSEM), and Geneva Finance Research Institute (GFRI); Bd du Pont d'Arve 40, 1211 Geneva 4, Switzerland; Rajna Gibson Brandon, rajna.gibson@unige.ch; Philipp Krueger, philipp.krueger@unige.ch; Peter S. Schmidt peter.steffen.schmidt@googlemail.com. We thank the editor (William Goetzmann), the co-editor (Luis Garcia-Feijoo), and two anonymous referees for their insightful comments. We also thank Stephen Brown and Oghuzan Karakas for valuable comments. The paper also benefited from comments received during the PRI conference in Paris in September 2019, the ECGI and University of Cambridge Conference "Engage or Divest? Strategies for responsible Investing" in Cambridge in October 2019, and a seminar audience at the University of Geneva. Supported by the Swiss National Science Foundation (SNSF) within the framework of the National Research Program "Sustainable Economy: resource-friendly, future-oriented, innovative" (NRP 73). Nadine Riand provided valuable research assistance in preparing the data used in this paper. The views expressed in this paper are solely those of the authors. We also remain responsible for any errors and omissions.

ESG rating disagreement and stock returns

Using ESG ratings from seven different data providers for a sample of S&P 500 firms between 2010 and 2017, we study the relation between ESG rating disagreement and stock returns. We find that stock returns are positively related to ESG rating disagreement, suggesting a risk premium for firms with higher ESG rating disagreement. The relation is primarily driven by disagreement about the environmental dimension. We discuss the practical implications of our findings for firms' equity cost of capital as well as for investment managers and asset owners who use ESG investment strategies.

Keywords: ESG ratings, disagreement, non-financial information, stock returns, equity cost of capital, sustainable finance

"There's so much disagreement about investing, and it's because nobody really knows."

Robert J. Shiller

Environmental, social, and governance (ESG) ratings are nowadays prominently featured in the financial press, regulatory and policy debates, academic studies, and are also a hot topic in investment practice. Such ratings also increasingly shape investment decisions of institutional investors representing trillions of dollars in assets under management (see Gibson, Glossner, Krueger, Matos, and Steffen 2021; GSIA 2016; USSIF 2020; PRI 2018). Recently, a lot of attention has been drawn to divergence of ratings issued by different ESG rating providers for the same firm. For example, in a Wall Street Journal article, Mackintosh (2018) points out that Tesla was rated highly by MSCI regarding environmental issues in 2018. In contrast, FTSE came to the opposite conclusion, rating Tesla poorly on environmental matters. Other news outlets, policy-oriented think tanks, and practitioner-oriented publications make similar observations (Wigglesworth 2018, Doyle 2018, Matos 2020).

In this paper, we pursue two objectives: first, we systematically analyze the level of disagreement about a firm's ESG rating to gain a better understanding of its magnitude and to determine whether ESG rating disagreement correlates with a firm's financial and accounting characteristics. Secondly, we study whether ESG rating disagreement has real consequences for firms and investors by examining if stock returns are related to ESG rating disagreement. While contemporaneous work attempts to explain the important question as to "why" ESG ratings disagree (e.g., Berg, Koelbel, and Rigobon 2020; Christensen, Serafeim, and Sikochi, 2021), this study is a first attempt at examining the fundamental issue of whether ESG rating

disagreement affects stock returns. Indeed, because of the prominent role that the equity cost of capital plays when financial analysts value firms and when CFOs decide on how to allocate capital expenditures, studying the role of ESG rating disagreement on stock returns, may be relevant to them. We furthermore discuss how ESG rating disagreement could impact the performance of equity investors who pursue sustainable investment strategies.

For the purpose of this paper, we collect and study ESG ratings from seven prominent ESG ratings providers for S&P 500 firms between 2010 and 2017. We use data from Asset 4 (now Refinitiv ESG), Sustainalytics (now Morningstar), Inrate, Bloomberg, FTSE, KLD (now MSCI), and MSCI IVA. We believe that our paper has the most comprehensive data coverage among current working papers which study the issue of ESG rating disagreement. For instance, Christensen, Serafeim, and Sikochi (2019) use data from three ratings providers only, while Berg, Koelbel and Rigobon (2020) do not focus on the time-series dimension. In contrast, we build a panel data set that uses seven different ESG ratings. Note also that the focus of the other papers is very different from ours in that they exclusively examine "why" ESG ratings disagree.

We start our analysis by documenting some very basic empirical facts about ESG rating disagreement in our sample of S&P 500 firms. We show, for example, that the average pairwise correlation between the ESG ratings of the seven rating providers is about 0.45. Surprisingly, the average pairwise correlation is lowest for the governance (0.16) and highest for the environmental dimension (0.46). Our analysis of pairwise correlations between ESG ratings from different providers also highlights more subtle patterns in disagreement, which go against the common belief of generalized ESG rating disagreement: for example, our analysis also shows that correlations between ratings from some providers are markedly higher, with the maximum correlation being about 0.75.

We then move on to studying whether disagreement varies along observable firm-level financial and accounting characteristics. Our paper provides evidence that disagreement tends to be higher for the largest firms in the S&P 500 (perhaps due to the complexity of such firms) and for firms that do not have a credit rating (potentially because the information environment for these firms being of lower quality). In contrast, more profitable firms tend to have lower ESG rating disagreement (perhaps because they are able to dedicate more resources to ESG policies and to their ESG disclosures). We also find that ESG rating disagreement is orthogonal to disagreement in the EPS forecasts issued by analysts. We further show that rating disagreement is generally more pronounced for firms that belong to the consumer durables and telecommunications industries, providing important insights for financial analysts that cover firms from these sectors.

In our main empirical analysis, we measure ESG rating disagreement using the standard deviation of the available ESG ratings from the seven different data providers for a given firm at a given point in time. We calculate the disagreement measures for the total ESG rating and separately for the E, S, and G dimensions (or "pillars"). We then relate monthly stock returns to our proxy of ESG rating disagreement, controlling for standard stock characteristics that are known to have predictive power in the cross-section of stock returns (e.g., size, momentum, quality).

We find that stock returns are positively related to ESG rating disagreement. Further tests show that the relation is driven mainly by disagreement about the environmental rating. In terms of economic magnitude, we estimate that an interquartile range increase in ESG rating disagreement is associated with an increase of 92 basis points in the annual cost of equity capital. Hence, ignoring differences in ESG rating disagreement in corporate valuation, could lead to sizeable mistakes when estimating the value of a firm's equity: assume, for instance, a

firm has perpetual annual Free Cash Flows to Equity (FCFE) of $100M per year, an expected growth rate of the FCFE of g = 0%, and a cost of equity capital r_E = 5 percent. An interquartile range increase in ESG rating disagreement combined with our empirical estimates would imply that the true cost of equity capital is 5.92 percent, suggesting over-estimating the value of the firm's equity by about $311M (or 18 percent) when ignoring ESG rating disagreement.[1]

Our findings of a positive relationship between ESG rating disagreement and stock returns can be rationalized with a standard asset pricing argument: higher ESG rating disagreement may be perceived as a source of uncertainty—in the spirit of Knightian uncertainty—that commands an uncertainty premium. This would explain why uncertainty-averse investors taking on such additional exposure wish to be compensated by higher expected returns. Consistent with this conjecture, we find in standard portfolio sorts that a portfolio which is long high disagreement stocks and short low disagreement stocks generates monthly returns of about 21 basis points (2.52 % on an annual basis) for disagreement about the overall ESG and the environmental rating. We find similar magnitudes when adjusting for well - known risk factors by using standard asset pricing models such as the Carhart (1997) model or the Fama and French (2015) five factor model.

While many papers (see, for instance, Friede, Busch, and Bassen (2015) and references therein) have studied the relation between stock returns and average ESG ratings, our paper is the first to systematically test whether the second moment of ESG ratings (i.e., disagreement about a firm's ESG performance) has consequences for stock returns. Another important contribution of our paper is to move beyond simply documenting that disagreement about non-financial information exists to shedding light on whether such ESG disagreement has real

[1] Valuation mistake = (Equity(no disagreement) - Equity(disagreement))/Equity(disagreement) – 1 = ([FCFE/(r_E - g)]/[FCFE/(r_E +0.0092-g)]= (100/(0.05))/(100/(0.05+0.0092))-1 = 0.184 (18.4%)

consequences for firms, analysts, and investors. We also contribute importantly to the debate on why ESG rating disagreement exists by providing evidence on the firms' financial and accounting characteristics that correlate with ESG rating disagreement and by identifying those industry sectors which are most prone to ESG rating disagreement.

Overall, our empirical results should help financial analysts, academics, institutional investors, financial advisors, policy-makers and regulators, and ultimately firms themselves to better understand that beyond the sustainability performance as captured by average ESG ratings, the dispersion of these ratings can also have an economically meaningful impact on stock returns and thus on firms' equity cost of capital.

1. Literature review

The use of ESG ratings in investment practice has increased considerably over the last two decades and has skyrocketed recently. In parallel, ESG ratings are now also commonly used in economics, management, and finance research[2]. Given the complexity of measuring a firm's non-financial or ESG performance, the validity and convergence of these ratings has been debated critically in the management literature.[3] Chatterji, Durand, Levine, and Touboul (2016), for instance, study the convergence of CSR ratings produced by six well-established information intermediaries. They document a lack of agreement across information intermediaries that comes mainly from two sources: the absence of both a common *theorization* and *commensurability*.[4] These findings point out that firms' and professional investors'

[2] see, for instance, Hong and Kostovetsky 2012; Krueger 2015; Lins et al. 2017; Liang and Renneboog 2017; Gibson, Krueger, and Mitali 2020; Dyck et al. 2019.

[3] See for instance Chatterji, Levine, and Toffel (2009), Bouten, Cho, Michelon, and Roberts (2017). or Delmas, Etzion, and Nairn-Birch (2013).

[4] The concept of a *common theorization* refers to the idea that raters (or information intermediaries) agree on a common definition of CSR. Absence of *commensurability* captures the idea that different raters do not use the same measures when quantifying the same feature.

sustainable financing and investment decisions are potentially tainted by the choice of their ratings' providers. The same applies to the conclusions stemming from existing empirical studies conducted by academics.

Given the heightened concern that ESG rating disagreement has generated in both practitioner circles and the financial press recently, the topic has also spurred significant academic interest. Most of the contemporaneous work aims at explaining the "why" that is the drivers of ESG rating disagreement. For instance, Christensen, Serafeim, and Sikochi (2021) focus on the role of disclosure as a determinant of ESG rating disagreement and find that more disclosure leads to higher disagreement.[5] In addition, they point out that the relationship between a firm's average ESG rating and ESG rating disagreement is non-linear. An important difference between their paper and ours is that they focus on explaining *why* disagreement exists, while we are more interested in examining the *consequences* arising from ESG rating disagreement, and specifically whether there are implications for stock returns, a firm's cost of capital and an investor's performance when investing sustainably. Berg, Koelbel, and Rigobon (2020) is another recent paper interested in explaining *why* disagreement exists. In their paper, the authors pursue a more granular approach and propose a decomposition of the sources of ESG rating disagreement. By subdividing the ESG ratings of six providers into finer categories, they identify three sources of ESG rating divergence. First, they highlight that raters use different categories, which can lead to disagreement. They refer to this as *scope divergence*. Secondly, they point out that raters measure identical categories differently, which they refer to as *measurement divergence*. Finally, they highlight *weight divergence*, which results from raters attaching different weights to the different categories when generating an aggregated ESG

[5] In a recent study Lopez-de-Silanes, McCahery, and Pudschedl (2019) find evidence that firms with good ESG scores may disclose more information.

rating. They find that most of the differences can be traced to measurement and scope divergence, while weight divergence seems to play a minor role. In addition, Berg, Koelbel, and Rigobon (2020) find a Rater Effect: ratings of one provider are positively correlated across different categories. The most important difference between these two papers and our study is that they focus on explaining *why* ratings disagree, while we are mainly interested in studying whether there are consequences from ESG ratings' disagreement, namely measurable effects on stock returns. In a recent paper Avramov, Chang, Lioui, and Tarelli (2021) examine the relationship between stock returns, ESG ratings, and ESG rating disagreement. Based on a theoretical model which highlights the interplay between the average ESG rating and ESG rating disagreement, Avramov et al. (2021) show that the average ESG rating is negatively associated with future stock performance only for low ESG disagreement stocks. In contrast to Avramov et al. (2021), the focus of our study is on ESG disagreement only.

Studying disagreement of ESG ratings is also reminiscent of the rich literature on heterogeneous beliefs in financial markets (see the discussion in Section 3.5). Many empirical studies have tested the relation between dispersion in beliefs and stock returns in a variety of settings. These studies typically use the dispersion in analyst earnings forecasts as a proxy for the extent to which a stock is subject to heterogeneous beliefs. For example, Diether, Malloy, and Scherbina (2002) document a significant and negative relation between heterogeneous beliefs and stock returns. Anderson, Ghysels, and Juergens (2005) obtain the opposite conclusion. They argue that disagreement about expected earnings per share is an additional priced risk factor and provide supportive empirical evidence for the pricing of this additional source of risk. We discuss another stream of the finance literature that would allow rationalizing the excess return for stocks with high ESG rating disagreement as a compensation for ESG information uncertainty in Section 3.5 (see, also, Viale, Garcia-Feijoo, and Giannetti, 2014).

2. Data

To test how stock returns are related to ESG rating disagreement, we construct a representative and homogeneous sample with available ESG ratings over the longest possible time period. We face the challenge that the availability of ESG data is restricted in both the cross-section and the time-series. This specific limitation is not unique to our setting but applies generally to research concerned with ESG. To use a sample as homogeneous as possible and to maximize the number of available ESG ratings per firm as well as the time-series dimension of the panel, we restrict ourselves to firms belonging to the S&P 500 and consider a sample period of eight years going from 2010 to 2017.

We use financial data from the Center for Research in Security Prices (CRSP) and accounting data from Standard & Poor's Compustat. We collect data from seven ESG data providers: (1) Asset 4 (now Refinitiv),[6] (2) Sustainalytics,[7] (3) Inrate, (4) Bloomberg, (5) FTSE, (6) KLD,[8] and (7) MSCI IVA.[9] According to a survey by Wong, Brackley and Petroy (2019) the most important and commonly used providers are Sustainalytics, MSCI, Bloomberg, and Asset 4. Later, we also conduct robustness analyses by restricting disagreement to the ratings from these four providers. In Appendix A, we provide further information on sample

[6] Asset 4 was acquired by Thomson Reuters in 2009, but the ESG data was made available under the old name of Asset 4. After the acquisition, the name was changed to *Thomson Reuters ESG Scores*. However, since the name Asset4 is widely known we use the old name for simplicity. Note that as of 2018, the ESG ratings data of Thomson Reuters is part of Refinitiv and now also known as Refinitiv ESG.

[7] After acquiring an about 40 percent stake in Sustainalytics in 2017, Morningstar purchased the remaining approximate 60 percent of Sustainalytics equity in 2020 (See https://bit.ly/3oXCgxM)

[8] The data from KLD originates from Kinder, Lydenberg, and Domini (KLD) Inc., which got acquired by Riskmetrics in 2009. In 2010, MSCI acquired Riskmetrics. Eccles, Lee, and Stroehle (2019) provide details on the history of KLD. We refer to these data as either KLD or MSCI KLD in this paper.

[9] The MSCI IVA dataset was initially created by Innovest, which was also acquired by Riskmetrics in 2009 before Riskmetrics got taken over by MSCI (see Eccles et al. (2019) for further details).

selection, dataset matching, variable definitions (See Appendix Table A.1), and summary statistics of the variables used in this study (see Appendix Table A.2).

Table 1 displays important features of these seven data providers. Column (1) shows the country origin for each provider. Three providers are US-based (Bloomberg, KLD, and MSCI IVA) whereas two providers have their origins in Switzerland (Asset 4 and Inrate). The other two providers can be traced back to origins in the Netherlands (Sustainalytics) and in the UK (FTSE).

[Table 1 about here.]

In Column (2), we show the rating scales used by each provider. Three providers apply a scale from 0 to 100 for their assessments (Asset 4, Sustainalytics, and Bloomberg), one provider uses a scale from 0 to 10 (MSCI IVA), one uses a scale from 0 to 5 (FTSE), and another provider a scale from 1 to 12 (Inrate), which is based on sustainability assessments ranging from D- to A+. Originally, MSCI KLD does not provide a genuine scale itself. However, many academic studies sum up KLD's strengths and concerns separately and scale both by the total number of strengths and concerns available. This course of action results in a scale from -1 to +1 (See, for example, Lins et al. 2017). Note that KLD also has strengths and concerns items for norms-oriented categories related to alcohol, military, firearms, gambling, nuclear, and tobacco, which we decide to ignore.

Because the rating scales differ not only in terms of their statistical support, but also in terms of the distribution across the statistical support, a simple re-scaling would not suffice to make the different ratings comparable. Therefore, we do the following to achieve comparability across rating providers: At each point in time, we sort all stocks according to the ratings of the

respective providers. We then calculate the individual rating specific percentile ranks and use these as adjusted scores. Using ranked measures is also more consistent with investment practice in which investors compare the ranked value of a given signal relative to the ranked values of the signal for other firms. When there are ties, we assign each company the average rank. We normalize these ranks between 0 and 1.

Column (4) shows the average number of sample stocks per year for which we observe an ESG rating from a given data provider. Sustainalytics, MSCI KLD, MSCI IVA and Bloomberg have on average the best coverage (about 460 stocks). Inrate, Asset 4 and FTSE have the least number of stocks on average with 434, 438 and 442, respectively. However, the average number of stocks for all providers is rather high with well above 400 and we therefore consider the sample as being representative for S&P 500 companies. Note however, that the Inrate and FTSE ratings are only available for a sub-period of the overall sample period. Therefore, we report in column (3) the time period for which data is available from a given provider.

The fifth column reports the pillar scores supplied by the providers as well as their rating styles. All providers supply a total ESG score, an environmental score, a social score, and a (corporate) governance score. In addition, Inrate also provides a labor score. Since the labor score captures a social topic, we use the average of the original social and the labor score as the social score. In Column (6) we highlight the rating styles employed by these various data-vendors that may partially explain disagreement. For instance, MSCI KLD and Inrate provide an absolute ESG rating, whereas providers such as Sustainalytics and MSCI provide Best-in-Class ratings. Also, some rating providers are more geared towards capturing ESG disclosure quality (e.g., Bloomberg).

3. Analysis

3.1 Descriptive statistics and correlations

Table 2 shows summary statistics and Pearson correlations between the ESG ratings from the seven different data providers. We display the results for the total rating and the three E, S, and G pillars in separate panels. The first three columns display descriptive statistics for the ranked ESG scores from the different providers. The subsequent columns display the pairwise cross correlations. We also show the average pairwise correlation between providers in the last row of each panel, which we calculate as the mean of the respective pairwise cross correlations (separately for the total rating and the E, S, and G pillars).[10]

[Table 2 about here.]

We first observe that the average pairwise correlation for the overall ESG ratings is 0.45, which is much lower than average correlations between credit ratings issued by Moody's and S&P. According to Berg et al. (2020) correlations among those two credit rating providers exceed 0.99. A point worth mentioning is the following: while there is a commonly held belief of generalized ESG rating disagreement, the analysis of pairwise correlations in Table 2 also highlights more subtle patterns in ESG disagreement in that the pairwise correlations between ratings from some providers can also be relatively high. For instance, the correlations between the total rating of Asset4 and Sustainalytics and Asset4 and Bloomberg are about 0.75 and that

[10] Each provider has a rather constant number of observations across the different scores they are issuing, with the exception of Bloomberg, which has substantially lower coverage for environmental ratings.

between Sustainalytics and Bloomberg about 0.69. This may be due to the fact that their rating styles are similar as was documented in Table 1.

Next, we move on to separately examine the average pairwise correlation between providers for the E, S, and G pillars (Panels B–D). These average pairwise correlations are generally lower than for the total rating, which is probably due to discrepancies in aggregation and weighting procedures across the three pillars. Surprisingly, the average correlation is lowest for the governance (0.16) and highest for the environmental ratings (0.46), but this can also be rationalized. While environmental issues can be increasingly measured and quantified (e.g., water usage, greenhouse gas emissions), the criteria applied to quantifying governance might differ across rating providers. In a similar spirit, the social rating is also likely to require more value judgements and is thus inherently more subjective, suggesting more disagreement (i.e., lower correlations). Our main argument for stating that the E rating is more objective and measurable is that we believe that there is not only more agreement on the issues which are important in the environmental dimension, but also more systematic regulation driven firm-level attempts of quantifying these dimensions. For instance, there is now a consensus that greenhouse gas emissions are an important dimension of a firm's environmental performance and GHG emissions are also increasingly measured. In a similar spirit firms, nowadays quantify water and electricity use. While measurement of emissions is certainly not without problems (scaling of emissions, missing emissions data/imputation, voluntary reporting), there is at least some basic guidance on how to measure these (see, for instance, the GHG protocol[11]).[12] In

[11] See http://www.ghgprotocol.org

[12] In addition, there is evidence from academics as well as from practitioners that environmental risk matters. Bolton and Kacperczyk (2021) find a sizeable risk premium for firms with high carbon emissions suggesting that investors do care about carbon risk. In a recent survey by Blackrock, 88% of the respondents place climate risk at the top of their portfolio concerns, "climate is king", emphasizing that Environmental risk is a major concern of their clients. (see https://bit.ly/3upXdlZ).

contrast, we think that when it comes to S and G there is (i) less agreement across providers on what the most important issues would be and (ii) a worse understanding of how to quantify these issues' real impacts.

[Figure 1 about here.]

In Figure 1, we look at whether the average pairwise correlations between ESG ratings vary at the industry-level. We plot average correlations across the seven ESG rating providers for each of the twelve Fama and French industries. There seems to be some industry heterogeneity when it comes to correlations between ESG ratings. Average correlations in the total ratings are lowest in the consumer durables and telecommunications sectors (see Subfigure 1(a)). The low average pairwise correlations in the total rating for the consumer durables industry seem to be driven by the low pairwise correlations in the environmental and governance rating (see subfigures 1(b) and 1(d)). The low average correlations of the telecommunications sector are due to the low correlations of the social ratings (see subfigure 1(c)). In contrast, ESG data providers seem to disagree the least (i.e., exhibit high average correlations) for the total rating in the business equipment and manufacturing sectors (see subfigure 1(a)). Another interesting observation is that rating providers also seem to disagree quite strongly (i.e., low correlations) about governance in the finance sector. These findings on industry-variation in firms' ESG ratings could help industry-focused financial analysts nuance their assessments and comparisons of firms in that analysts are most of the time confronted with metrics coming from various data-providers.

3.2 Determinants of ESG rating disagreement

While we do not see our main contribution as studying determinants of ESG rating disagreement, the comprehensiveness of our sample (in terms of rating providers covered) as well as the representativeness of the sample-firms (S&P 500 firms) allow us to also contribute to this literature. Hence, to add to the existing literature that traces the origins to disagreement in scopes, ratings methodologies and ESG firms' disclosure policies, we are the first to examine whether ESG rating disagreement also correlates with observable firm-level financial and accounting characteristics. We use the standard deviation of ESG ratings available for a given firm at a given point in time as the dependent variable. We calculate this measure for the total rating, but also separately for the E, S, and G pillars. We explore how variables falling in one of the following five categories correlate with disagreement: (i) Balance sheet related, (ii) Industry related, (iii) Investor transparency, (iv) Valuation, and (v) Price.[13]

We use pooled panel regressions in which the rating disagreement measures serve as dependent variables. We also include industry-month fixed effects. Standard errors are double clustered at the firm and month-level. In Column (1) of Table 3 we display the regression results for disagreement about the total rating. Columns (2)—(4) display the results separately for disagreement about the E, S, and G pillars.

[Table 3 about here.]

[13] (i) Balance sheet related: Tangibility (*TAN*), current ratio (*CR*), leverage (*LEV*), gross profitability (*GP*) (Novy-Marx 2013); (ii) Industry related: Industry concentration measured by the Herfindahl-Hirschman index (*HHI*) based on book equity, multi-segment (*MSEG*); (iii) Investor transparency: Missing credit rating (*NCR*), institutional ownership (*IO*), number of analysts (*NoA*), the dispersion of analyst forecasts of the firm's one year ahead earnings forecasts (*StdDev*) (Diether et al. 2002); (iv) Valuation: Book-to-market (*BM*) (Fama and French 1995); (v) Price: market cap (*ME*) (Banz 1981), momentum (*MOM*) (Jegadeesh and Titman 1993), and total volatility (*TVOL*) (Ang, Hodrick, Xing, and Zhang 2006).

Essentially, three financial variables play a role in explaining ratings disagreement. First, more profitable firms (in terms of gross profitability - *GP*) are subject to lower ESG rating disagreement (see columns 1 and 2). Secondly, firms without a credit rating *(NCR)* exhibit higher ESG rating disagreement (see columns 1 and 3), as do larger firms (see columns 3 and 4).[14] These results seem intuitive: Profitable firms may be viewed less critical by ESG analysts, perhaps because they have more resources available to shape and disclose their ESG policies. In contrast, firms without a credit rating are subject to a less transparent information environment, making their assessment in terms of ESG more difficult. Also, larger firms might be more diversified and complex and are also analyzed more thoroughly by ESG data providers, explaining why they exhibit higher ratings disagreement. Finally, it is interesting to note that ESG rating disagreement is orthogonal to disagreement about EPS earnings (variable *StdDev* in Table 3).

In addition, three other variables seem to affect disagreement about individual pillars of the ESG rating. For instance, Tangibility *(TAN)* plays a specific role in that firms with more tangible assets tend to have lower disagreement in their environmental rating. Again, this seems intuitive given that firms with more tangible assets are also likely to have a more negative impact on the environment (e.g., higher greenhouse gas emissions) and thus a potentially more easily measurable environmental rating. Moreover, higher levels in institutional ownership *(IO)* are also associated with higher disagreement in the environmental rating. Firms with a high book-to-market ratio *(BM)* display higher disagreement in their social rating.

[14] The reader might wonder why S&P 500 firms do not have a credit rating. In general, firms without a credit rating do not seem to be exceptional. For example, in a sample of 12,312 firms, Avramov, Chordia, Jostova, and Philipov (2009) report that 9,051 firms do not have a credit rating. In our sample 194 out of a total of 553 firms do not have a credit rating for at least one month.

3.3 ESG rating disagreement and stock returns

We now turn to our main research question and examine the relationship between stock returns and ESG rating disagreement. As in the previous section, we rely on pooled panel regressions with standard errors double clustered at the firm and month-level. We use monthly stock returns as the dependent variable in the regressions. Besides our main disagreement related explanatory variables, which we measure as the standard deviation of ratings available for a given firm at monthly intervals and denote as *Disp*, we include industry-month fixed effects and also control for standard characteristics that have been found to explain the cross section of stock returns. We control for market capitalization (Banz 1981), book-to-market (Fama and French 1995), gross profitability (Novy-Marx 2013), momentum (Jegadeesh and Titman 1993), the dispersion of analyst forecasts of the firm's one year ahead earnings forecasts (Diether et al. 2002), the firm's beta (Frazzini and Pedersen 2014) and total volatility (Ang et al. 2006). Conceptually, pooled panel regressions with industry-month fixed effects are similar to Fama and MacBeth (1973) type regressions with industry dummies. Given that our sample period is relatively short and also because ESG rating disagreement varies across sectors (see Figure 1), it is important to control for differences at the industry-level.

[Table 4 about here.]

The coefficient estimates for the main explanatory variables *Disp* and the control variables are displayed in Table 4. We also show *t*-statistics based on double clustered standard errors (in parentheses). Regression results for disagreement about the overall ESG rating are tabulated in Column (1). The results for disagreement about the E, S, and G pillars are separately in columns (2) to (4).

We observe a positive and significant coefficient estimate for the total ESG rating disagreement proxy suggesting that firms with higher disagreement tend to have higher stock returns. In the regression of Column (1), we estimate a coefficient of about 0.7 for disagreement about the total ESG rating. In terms of economic magnitude, consider a firm moving from the 1st quartile (0.14) to the 3rd quartile (0.25) of the ESG rating disagreement distribution. Such a move would imply an increase of about 92 basis points in annual stock returns (= (0.25-0.14)*0.7*12), an estimate which seems plausible in terms of magnitude.

In columns (2) to (4), we examine the effect of rating disagreement for the E, S, and G pillar separately. In Column 2, we find a coefficient estimate for disagreement about the environmental rating of about 1.0, which is both economically larger and more significant (*t*-statistic=2.38) when compared to the estimate for disagreement about the total rating. The coefficients for disagreement about the social rating (see Column 3) is positive (0.6) but not significant at conventional levels (*t*-statistic=1.51). The coefficient estimate for disagreement about the governance rating (see Column 4) rating is small (-0.06) and insignificant (*t*-statistic=-0.136). We conclude that disagreement about the environmental rating primarily drives our results, with an interquartile range increase in disagreement about the environmental rating resulting in an increase of about 132 basis points (=(0.25-0.14)*1*12)) in the annual cost of equity capital of a firm.

Given that ratings from Inrate and FTSE are not available for the entire sample period, we re-estimate the regressions using a disagreement measure based only on Asset 4, Sustainalytics, Bloomberg, MSCI KLD and MSCI IVA. The results are reported in Appendix Table B.1 and continue to show a strong positive relation between ESG rating disagreement and stock returns. In fact, the results in these regressions are stronger both statistically and economically speaking, when looking at disagreement about the E and S pillar separately. The *t*-statistic for

disagreement about the environmental rating increases to 3.01, and disagreement about the social rating is now also marginally significant (t-statistic=1.72).

The reader may wonder why the control variable in the return regressions of Table 4 do not turn out to be significant. This might have a variety of reasons. First, the control variables are known return predictors and there is evidence of lower post-publication return predictability (McLean and Pontiff 2016), especially in the U.S (Jacobs and Müller 2020). Other research also shows that return predictability fell sharply post 2003 (Green, Hand, and Zhang 2017). In a similar spirit Chordia, Subrahmanyam, Tong (2014) show that capital market anomalies have attenuated in recent periods, coinciding with periods that have been accompanied by significant liquidity increases. Given the combined findings of these papers and noting that we are studying a sample of highly liquid S&P 500 firms since 2010, it is not surprising to observe that the control variables are insignificant.

Overall, we conclude that higher stock returns for firms with higher ESG rating disagreement is consistent with the view that risk averse investors perceive more uncertainty about the ESG performance of a given firm as an additional source of risk (or uncertainty) that commands a separate premium.

3.4 Portfolio sorts on ESG rating disagreement

In the previous sub-section we have documented a positive relationship between ESG rating disagreement and stock returns. To evaluate whether ESG rating disagreement provides a profitable signal for investing, we now implement portfolios sorts based on ESG rating disagreement.

In the portfolio sorts, we use industry-adjusted disagreement as the sorting variable. We do so primarily to rule out the possibility that our results are driven by industry effects. As Figure

1 has shown, average ESG rating correlations and thus disagreement exhibit important industry variation and we want to ensure that the portfolio composition in the sorts are not biased by these industry differences in ESG rating disagreement. Note that in the return regressions of Table 4, we also controlled for industries by including industry-month fixed-effects, implying that our insights and results are driven by differences in disagreement *within* and not *between* industries.

More specifically, we calculate the industry-adjusted ESG rating disagreement for a given firm in a given month by simply de-meaning stock-level ESG rating disagreement using the average ESG rating disagreement in the firm's Fama-French 12 industry in a given month. We denote this variable as *Disp_adj* and we sort all stocks in the sample into five quintile groups based on their industry-adjusted ESG rating disagreement. The results are shown in Table 5.

[Table 5 about here.]

For each of the four rating dimensions (Total, Environmental, Social, and Governance) we report results for the equally-weighted portfolios of stocks with the lowest industry-adjusted rating disagreement (low *Disp_adj*, first quintile), the highest industry-adjusted rating disagreement (high *Disp_adj*; fifth quintile), both in excess of the risk-free rate, as well as the long-short portfolio between high and low disagreement stocks (H-L *Disp_adj*). We report mean monthly returns (*Ret*), median number of firms for the low and high disagreement portfolios (*N*), standard deviations (*Stddev*), Sharpe-Ratios (*SR*), as well as alphas computed from four different factor models (CAPM, Fama-French 3 factor model (*FF3*), Carhart 4 factor model (*Car*), and the Fama-French 5 factor model (*FF5*)). We observe in Panel A of Table 5 that for the total rating, the equally weighted long-short portfolio between high and low disagreement stocks generates a mean raw monthly return of 21 basis points (with a *t*-statistic

of 2.2). Factor model alphas are of similar magnitude with average monthly alphas ranging between 23 and 27 basis points (and *t*-statistics between 2.0 and 2.7). Furthermore, we observe that a long-only strategy of high disagreement stocks generates a raw monthly mean return of 134 basis points (with a *t*-statistic of 3.5). The alphas of the long-only strategy, ranging between 14 and 21 basis points (with *t*-statistics between 1.6 and 2.4) are somewhat lower.

For the Environmental dimension (see Panel B, Table 5), we observe similar results. The long-short portfolio generates a mean monthly raw return of 21 basis points (with a *t*-statistic of 2.0). The factor model alphas range between 21 and 26 basis points (with *t*-statistics between 2.0 and 2.5). The long-only strategy also shows a high average monthly raw return of about 140 basis points (*t*-statistic of 3.5), and also high average monthly factor alphas, ranging from 16 to 22 basis points (with *t*-statistics between 1.8 and 2.3). For disagreement about the social and governance ratings, we do not obtain any significant raw returns or alphas from the long-short strategies. Taken together, the insights from the portfolio sorts in Table 5 are consistent with the evidence presented in the regression analysis of Table 4 and emphasize the prominent role played by the environmental rating disagreement on stock returns.

One important question is whether these strategies are implementable and survive accounting for trading costs. To get a realistic impression as to whether this is the case, we follow the approach advocated by Novy-Marx and Velikov (2016). Novy-Marx and Velikov (2016) utilize the effective bid-ask spread measure of Hasbrouck (2009) to proxy for trading costs. Hasbrouck (2009) suggests a Bayesian Gibbs sampler approach to the Roll (1984) model of price dynamics. Hasbrouck (2009) shows that his estimate of the average effective cost is comparable to estimates from high-frequency trade and quote (TAQ) data and reports a 0.965 Pearson-correlation of his measure with the TAQ value. Novy-Marx and Velikov (2016) note that this measure does not account for the price impact of large trades, but nevertheless consider

it as an appropriate trading cost measure (for details see Novy-Marx and Velikov (2016), p. 108).

To calculate the actual trading costs for each portfolio, we subtract trading costs from portfolio returns each time a given stock is introduced into a portfolio or withdrawn from it.

[Table 6 about here.]

The results of the long–short strategies after accounting for transaction costs are displayed in Table 6. We observe that the raw returns, Sharpe ratios and alphas are somehow diminished, but still sizable. For example, for the total rating disagreement, the monthly raw return of the long-short portfolio between high and low disagreement stocks drops from 21 basis points (see Table 5) to 18 basis points (Table 6). The *t*-statistic drops from 2.2 to 1.9 (which still implies significance at the 10 percent level). The monthly alphas range from 19 to 23 basis points (with *t*-statistics of 1.8 to 2.4). For the Environmental rating disagreement, we observe significant monthly alphas for the CAPM and for the Fama-French 3 factor and the Carhart 4 factor models. These alphas range between 18 and 22 basis points (with t-statistics between 1.9 and 2.2). However, the long-short raw returns and the alphas from the Fama-French 5 factor model are slightly below the 10 percent threshold with t-statistics of about 1.6 in both cases (and mean returns of about 17 basis points). As in the case without transaction costs, the long-short portfolios for the social and the governance ratings' disagreement are displaying insignificant raw returns and risk-adjusted alphas.

3.5 Possible theoretical explanations

A potential way of rationalizing the relation between ESG ratings disagreement and stock returns may be found in the literature on heterogeneous beliefs in financial markets. Theoretical models of heterogeneous beliefs (e.g. Atmaz and Basak, 2018) provide some predictions for empirical studies. It should be noted, however, that these theories are based on beliefs about factors that affect the returns of firms (e.g., consensus EPS forecasts) and not factors that may or may not have risk and return implications like ESG ratings. Nonetheless, we still think that it is useful to discuss our findings in the light of this literature without attempting to test a specific theory in our setting.

In an important paper, Atmaz and Basak (2018, p. 1241) argue that "…dispersion represents an additional risk for investors and therefore investors demand a higher return to hold the stock when dispersion is higher." In their setting, this impact may however be attenuated (reinforced) in the presence of investors' excessive optimism (pessimism). Along the same lines, Anderson, Ghysels and Juergens (2005) find that adding empirical factors based on analysts EPS forecast dispersion, enhances explanatory power in explaining S&P 500 excess returns after accounting for standard market risk factors. Translated to our context, that would imply that ESG rating disagreement risk is priced in expected stock returns in addition to the market risk exposure of stocks. In other words, our findings of a positive relation between stock returns and ESG rating disagreement is in line with the risked-based explanation for analysts EPS disagreement empirically documented by Anderson, Ghysels and Juergens (2005). According to this risk-based explanation, higher total (or environmental) rating disagreement would imply more uncertainty about the ESG performance of a given firm and thus would be perceived as a separate source of risk that commands a risk premium if investors are risk averse. We find supportive empirical evidence that this is the case for environmental and total rating

disagreement and thereby add to this stream of literature by documenting the effects of heterogeneity in beliefs about non–financial information.

Another possible explanation for our findings is that ESG rating disagreement is a proxy for ESG uncertainty and therefore captures a specific form of (Knightian) uncertainty. Whereas risk is associated with the uncertain outcome of a known probability distribution of returns, ambiguity (or Knightian uncertainty) is associated with uncertainty regarding the probability distribution itself (see for example, Viale, Garcia-Feijoo and Giannetti, 2014). Theoretical papers advocate to use a two-factor model with one factor as a proxy for risk and the other factor as a proxy for uncertainty (see in particular, Kogan and Wang, 2003). Anderson, Gyhsels and Juergens (2009) estimate a two-factor model with a factor-proxy for uncertainty based on professional forecasters' disagreement. Viale, Garcia-Feijoo and Giannetti (2014) further show that uncertainty (or ambiguity) is priced in the cross -section of stock returns, but not subsumed by standard risk factors. Hence, in the context of our setting, a possible explanation for the positive relationship between ESG rating disagreement and stock returns is that our ESG rating disagreement measure is a proxy for uncertainty regarding ESG information and in particular about environmental information.

3.6 Limitations

Our study bears several limitations. First, the fact that we focus on the S&P 500 universe leaves it open as to whether our results can be transposed to other stock markets located outside of the U.S. It is also unclear whether and how the conclusions hold in larger cross-sections of stocks. Secondly, to encompass as many data-providers as possible, we had to work with a very limited time period that extends from 2010 until 2017. Thus, the power of our tests might be an issue.

This issue can perhaps explain why we do not find any significant effects for disagreement about the S and G ratings. Third, some of the data-vendors may have changed their rating methods during our sample period that may create additional biases when measuring ESG rating disagreement. Finally, we have only focused on stocks but it may be interesting to explore whether our results can be transposed to fixed-income securities that have ESG ratings.

3.7 ESG rating disagreement and equity risk

For completeness, we examine in this last subsection whether ESG rating disagreement affects standard equity risk measures. An examination of the relation between equity risk and ESG disagreement seems natural given that prior research highlights a robustly negative relation between stock-level risk and the quality of a firm's ESG rating (Dunn, Fitzgibbons, and Pomorski, 2018; Hoepner, Oikonomou, Sautner, Starks, and Zhou, 2019; Gibson et al. 2020). Hence, we regress standard risk measures on our ESG rating disagreement proxies and a set of common control variables (see Table 7).[15] We focus on the three risk measures *Total Volatility*, *Idiosyncratic Volatility*, as well as *Beta*, and, following prior studies, also analyze whether downside risk as measured by Lower Partial Moment *(LPM)* is related to ESG rating disagreement (see Hoepner et al (2019) for details on the relation between ESG policies and LPM).

[Table 7 about here.]

[15] We include industry-month fixed effects and control for standard characteristics (not reported) that have been found to explain volatility (see for example Dennis and Stickland, 2004), namely market capitalization (Banz 1981), leverage, business segment (dummy variable, which is one for multi-segment firms), percentage of institutional ownership, the ratio of mutual fund ownership to total institutional ownership, and turnover.

We observe that some risk measures are positively related to ESG rating disagreement, in particular to disagreement about the social dimension. For example, *Total Volatility*, *Idiosyncratic volatility,* and *LPM* all appear to be positively related to disagreement about the social rating (see Column 3). In addition, there is some evidence that idiosyncratic and downside risk are related to total rating disagreement. We believe that in light of the second theoretical explanation provided in Section 3.5 ("Knightian uncertainty"), it is not surprising to observe that ESG rating disagreement is only weakly related to standard risk measures. The uncertainty-based explanation is also consistent with the fact that ESG rating disagreement has very little explanatory power in explaining standard risk measures. Therefore, our results suggest that ESG information uncertainty offers explanatory power beyond traditional risk factors.

4. Conclusion

Recently, the issue of ESG rating disagreement has received considerable attention from the financial press, and from practitioner and policy-making circles. In addition, ESG rating disagreement has important implications for the generalization of academic research findings and is creating challenges for asset managers in their efforts to implement ESG investment strategies. We provide a first step towards a better understanding of the real consequences of ESG rating disagreement by studying its impact on stock returns. More specifically, we find that stock returns are positively related to ESG rating disagreement and in particular to the environmental rating disagreement. We do not assume per se that the environmental pillar is more important than the social and the governance pillars, but rather find this as a result of the analysis. In addition, academics (see Bolton and Kacperczyk, 2021) as well as practitioners (see Blackrock survey 2020) emphasized recently that environmental risks are important, thus this

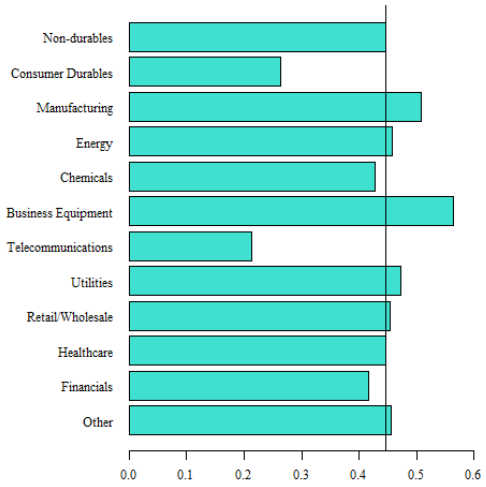
may point to the fact that the disagreement about the environmental rating dimension is the only one priced so far by investors with ESG preferences.

Our results have important practical consequences. First, our analysis shows that financial analysts who value the equity of firms should incorporate the effects of ESG rating disagreement and adjust their equity cost of capital estimates upwards. Second, CFOs deciding about the allocation of capital expenditures should also consider ESG rating disagreement in their capital budgeting decisions as it raises the investment threshold for firms subject to high total (and environmental) rating disagreement. Thirdly, our evidence that ESG rating disagreement varies across industries is also an important insight for financial analysts, who often focus on specific industries.

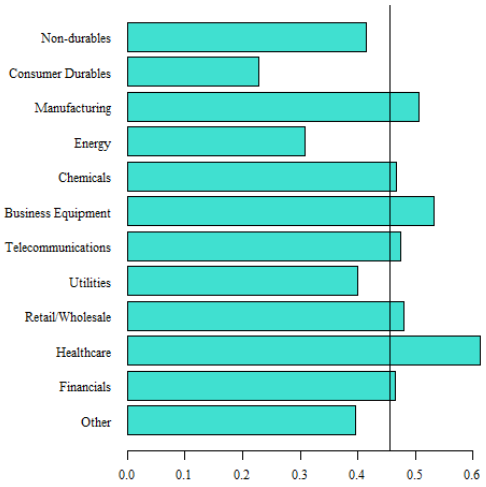
Finally, our analysis also has important implications for asset owners and investment managers who implement responsible investment strategies. In particular, two strategies are currently very popular in the responsible investment landscape, namely screening and ESG integration. If asset managers and investment managers wish to optimize financial performance while investing responsibly, they should care about ESG rating disagreement and its impact on stock returns. Indeed, our results suggest that with positive (negative) screening they should buy (sell) primarily those stocks that, for a given high (low) ESG rating, command the lowest (highest) level of ESG disagreement. This should allow positive (negative) screeners to mitigate the adverse impact of ESG rating disagreement on their buy (sell) orders expected future returns. Similarly, an ESG integration strategy may fail on delivering its financial promises if it does not search simultaneously for those stocks that have superior ESG ratings and embed the lowest level of ESG rating disagreement within an industry. Indeed, controlling for a low level of ESG rating disagreement will allow investors who integrate ESG criteria in their stock selection process to avoid a subsequent unintended stock price decline.

Figure 1: Average correlations by Fama and French 12 industry

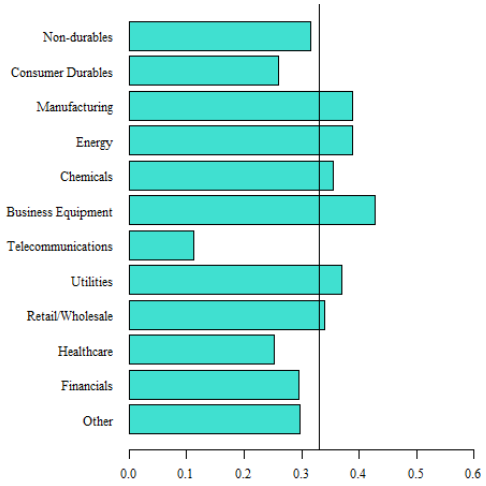
(a) Total rating



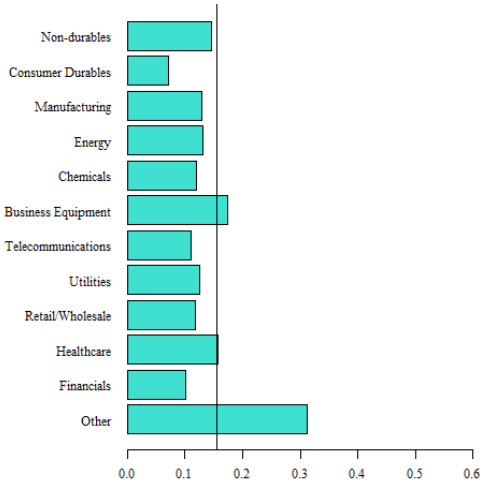
(b) Environmental pillar



(c) Social pillar



(d) Governance pillar



Note: This figure plots average pairwise Pearson correlations between the ratings of the seven different ESG data providers for each of the twelve Fama and French industries. We report average correlations for the total rating in subfigure 1a and the respective ESG components in subfigures 1b, 1c, and 1d. The vertical line represents the average correlation across all industries.

Table 1: ESG data providers

Data provider	**Origin**	**Rating scale**	**Period covered**	**Number of stocks (sample)**	**Pillars**	**Rating Style**
	(1)	(2)	(3)	(4)	(5)	(6)
Asset4	CH	0 – 100	2010/01-2017/12	438	E, S, G, Total	Disclosure oriented
Sustainalytics	NL	0 – 100	2010/01-2017/12	459	E, S, G, Total	Best in Class
Inrate	CH	1 – 12	2013/01-2017/12	434	E, L, S, G, Total	Absolute ESG Rating
Bloomberg	US	0 – 100	2010/01-2017/12	463	E, S, G, Total	Disclosure oriented
FTSE	UK	0 – 5	2014/10-2017/12	442	E, S, G, Total	Best in Class
MSCI KLD	US	-1 – +1	2010/01-2017/12	468	E, S, G, Total	Absolute ESG rating
MSCI IVA	US	0 – 10	2010/01-2017/12	456	E, S, G, Total	Best in Class

Note: This table provides an overview of the ESG data providers which we use in this study. We list the name of the respective data provider (Data provider), the country in which the data provider has its origins (Origin), the rating scale used by the respective data provider (Rating scale), the average number of stocks per year in the sample for the total rating of each provider (Number of stocks (sample)), and the data dimensions (e.g., environmental, social, and governance) that are available from each provider. We refer to these data dimensions as Pillars.

Table 2: Descriptive statistics and correlations

	N	Mean	StdDev	Pearson correlations					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(8)
				Asset4	Sust.	Inrate	Bloom.	FTSE	KLD
Panel A: Total Pillar									
Asset4	42'087	0.501	0.289						
Sustainalytics	44'078	0.501	0.289	0.752					
Inrate	26'037	0.501	0.284	0.233	0.303				
Bloomberg	44'464	0.501	0.289	0.750	0.693	0.124			
FTSE	17'220	0.501	0.288	0.568	0.614	0.267	0.586		
KLD	44'951	0.501	0.288	0.524	0.559	0.292	0.477	0.488	
MSCI IVA	43'775	0.501	0.289	0.396	0.434	0.318	0.303	0.266	0.439
Average correlation						**0.447**			
Panel B: Environmental Pillar									
Asset4	42'019	0.501	0.289						
Sustainalytics	44'020	0.501	0.289	0.706					
Inrate	26'036	0.501	0.286	0.305	0.487				
Bloomberg	37'624	0.501	0.289	0.647	0.557	0.206			
FTSE	17'220	0.501	0.288	0.654	0.678	0.368	0.607		
KLD	44'669	0.501	0.280	0.575	0.609	0.422	0.431	0.581	
MSCI IVA	43'580	0.501	0.289	0.233	0.352	0.404	0.187	0.239	0.312
Average correlation						**0.455**			
Panel C: Social Pillar									
Asset4	42'087	0.501	0.289						
Sustainalytics	44'078	0.501	0.289	0.617					
Inrate	26'037	0.501	0.288	0.133	0.143				
Bloomberg	44'364	0.501	0.288	0.685	0.527	0.062			
FTSE	17'220	0.501	0.288	0.637	0.501	0.106	0.560		
KLD	44'951	0.501	0.288	0.367	0.391	0.129	0.276	0.271	
MSCI IVA	43'775	0.501	0.289	0.266	0.303	0.236	0.202	0.191	0.337
Average correlation						**0.330**			
Panel D: Governance Pillar									
Asset4	42'087	0.501	0.289						
Sustainalytics	44'078	0.501	0.289	0.331					
Inrate	26'037	0.501	0.283	0.297	0.401				
Bloomberg	44'464	0.501	0.282	0.432	0.327	0.344			
FTSE	17'220	0.501	0.288	0.027	0.160	-0.029	-0.027		
KLD	44'951	0.501	0.248	0.104	0.089	0.081	0.153	-0.065	
MSCI IVA	43'775	0.501	0.288	0.132	0.135	0.145	0.060	0.023	0.133
Average correlation						**0.155**			

Note: This table shows summary statistics and Pearson correlations between the ratings of the seven different data providers. The results are displayed in separate panels for the Total rating and the E, S, and G pillars. The first three columns show the descriptive statistics of the different ESG providers' ranked scores (number of observations (*N*), mean (*Mean)*, and standard

deviation (*StdDev*)). The following columns display the pairwise cross-correlations. We also display the average pairwise correlation between providers in the last row of each panel.

Table 3: Determinants of ESG rating disagreement

Dependent variable:	ESG rating disagreement			
	(1)	(2)	(3)	(4)
Pillars:	Total	Environmental	Social	Governance
Balance Sheet related				
TAN	-0.013	**-0.030**	-0.010	-0.011
	(-0.909)	**(-2.027)**	(-0.673)	(-0.859)
CR	0.012	-0.001	-0.001	0.010
	(1.064)	(-0.081)	(-0.114)	(1.230)
LEV	-0.015	-0.012	0.006	-0.005
	(-1.464)	(-1.163)	(0.657)	(-0.594)
GP	**-0.028**	**-0.027**	-0.021	-0.010
	(-1.954)	**(-1.955)**	(-1.610)	(-0.786)
Industry				
HHI	**0.029**	0.021	0.012	-0.008
	(2.183)	(1.593)	(0.911)	(-0.775)
MSEG	0.001	-0.002	0.002	0.002
	(0.085)	(-0.293)	(0.371)	(0.416)
Investor Transparency				
NCR	**0.015**	-0.010	**0.020**	**-0.012**
	(1.678)	(-1.208)	**(2.713)**	**(-1.851)**
IO	0.014	**0.015**	0.010	-0.004
	(1.491)	**(1.794)**	(1.073)	(-0.551)
NoA	-0.008	-0.011	-0.004	0.000
	(-0.906)	(-1.183)	(-0.446)	(0.035)
StdDev	0.005	0.006	0.013	0.005
	(0.581)	(0.887)	(1.579)	(0.646)
Valuation				
BM	0.010	-0.006	**0.022**	0.005
	(0.793)	(-0.524)	**(1.936)**	(0.495)
Price				
ME	0.018	0.012	**0.033**	**0.019**
	(1.514)	(1.085)	**(2.854)**	**(1.905)**
Momentum	0.003	-0.001	0.003	-0.003
	(0.573)	(-0.260)	(0.590)	(-0.740)
TVOL	-0.003	-0.005	-0.001	-0.008
	(-0.272)	(-0.541)	(-0.095)	(-0.948)
Industry * Month FE	Yes	Yes	Yes	Yes
N	35,139	34,902	35,139	35,139
Adjusted R^2	0.059	0.057	0.047	0.052

Note: This table displays the results of pooled panel regressions in which ESG rating disagreement is regressed on observable firm characteristics. We measure disagreement as the standard deviation of all firm level ratings available for a given firm at a given point in time. We use disagreement about the total rating (Column 1) and the E, S, and G pillars separately (columns 2—4). The explanatory variables are the following: tangibility (*TAN*), current ratio (*CR*), leverage (*LEV*), gross profitability (*GP*) (Novy-Marx 2013), Herfindahl-Hirschman index (*HHI*), multisegment (*MSEG*), missing credit rating (*NCR*), institutional ownership (*IO*), number of analysts (*NoA*), the dispersion of analyst forecasts of the firm's one year ahead earnings forecasts

(*StdDev*) (Diether et al. 2002), book-to-market (*BM*) (Fama and French 1995), market cap (*ME*) (Banz 1981), momentum (*MOM*) (Jegadeesh and Titman 1993), and total volatility (*TVOL*) (Ang et al. 2006). We also include industry-month fixed effects. *t*-statistics based on double clustered standard errors (month and firm) are reported in parentheses.

Table 4: Stock returns and ESG rating disagreement

Dependent Variable:	Returns			
	(1)	(2)	(3)	(4)
Pillars:	Total	Environmental	Social	Governance
Disp	**0.698**	**1.012**	0.630	-0.059
	(1.995)	**(2.375)**	(1.515)	(-0.136)
ME	-0.134	-0.128	-0.135	-0.119
	(-0.726)	(-0.686)	(-0.725)	(-0.635)
BM	0.216	0.233	0.218	0.226
	(0.973)	(1.048)	(0.984)	(1.021)
GP	0.231	0.246	0.234	0.222
	(0.772)	(0.825)	(0.781)	(0.745)
Momentum	0.360	0.348	0.359	0.361
	(1.114)	(1.064)	(1.109)	(1.120)
StdEPS	-0.232	-0.193	-0.233	-0.228
	(-1.337)	(-1.116)	(-1.344)	(-1.313)
Beta	0.165	0.188	0.168	0.158
	(0.330)	(0.376)	(0.336)	(0.316)
TVol	-0.081	-0.115	-0.084	-0.073
	(-0.227)	(-0.328)	(-0.237)	(-0.206)
Industry * Month FE	Yes	Yes	Yes	Yes
N	42,058	41,786	42,058	42,058
Adjusted R^2	0.347	0.348	0.347	0.347

Note: This table displays the results of pooled panel regressions of monthly stock returns on ESG rating disagreement. The first row reports the results for disagreement in the total ESG rating and the rows (2) – (4) report the results for the E, S, and G pillar separately. The dependent variable *Returns* is the firm's monthly stock return. We measure ESG rating disagreement by the standard deviation of ratings available for a given firm at a given point in time (*Disp*). We also include industry-month fixed effects and control for standard characteristics that have been found to explain stock returns, namely market capitalization (Banz 1981), book-to-market (Fama and French 1995), gross profitability (Novy-Marx 2013), momentum (Jegadeesh and Titman 1993), the dispersion of analyst forecasts of the firm's one year ahead earnings forecasts (Diether et al. 2002), the firm's beta (Frazzini and Pedersen 2014), and total volatility (Ang et al. 2006). *t*-statistics based on double clustered standard errors (month and firm) are reported in parentheses.

Table 5: Portfolio sorts on industry-adjusted ESG rating disagreement

	(1) Ret	(2) N	(3) Stddev	(4) SR	(5) CAPM	(6) FF3	(7) Car	(8) FF5
					Panel A: Total			
Low Disp_adj	**1.124**	94	1.124	0.295	-0.088	-0.068	-0.055	-0.074
	(2.891)				(-1.260)	(-0.970)	(-0.794)	(-1.049)
High Disp_adj	**1.336**	94	1.336	0.355	0.144	**0.164**	**0.211**	0.165
	(3.474)				(1.561)	**(1.792)**	**(2.425)**	(1.611)
H-L Disp_adj	**0.212**		0.212	0.225	**0.232**	**0.232**	**0.267**	**0.239**
	(2.209)				**(2.151)**	**(2.192)**	**(2.664)**	**(2.044)**
					Panel B: Environmental			
Low Disp_adj	**1.186**	93	1.186	0.292	-0.091	-0.043	-0.016	-0.036
	(2.866)				(-0.902)	(-0.500)	(-0.187)	(-0.455)
High Disp_adj	**1.397**	93	1.397	0.358	**0.163**	**0.183**	**0.212**	**0.178**
	(3.504)				**(1.848)**	**(2.040)**	**(2.274)**	**(1.982)**
H-L Disp_adj	**0.211**		0.211	0.205	**0.254**	**0.226**	**0.228**	**0.213**
	(2.010)				**(2.482)**	**(2.226)**	**(2.322)**	**(2.005)**
					Panel C: Social			
Low Disp_adj	**1.283**	93	1.283	0.334	0.071	0.121	**0.139**	**0.116**
	(3.273)				(0.865)	(1.734)	**(1.943)**	**(1.950)**
High Disp_adj	**1.330**	94	1.330	0.333	0.067	0.096	**0.134**	0.081
	(3.262)				(0.658)	(1.077)	**(1.748)**	(0.998)
H-L Disp_adj	0.047		0.047	0.049	-0.004	-0.024	-0.004	-0.035
	(0.480)				(-0.058)	(-0.338)	(-0.073)	(-0.479)
					Panel D: Governance			
Low Disp_adj	**1.246**	94	1.246	0.320	0.013	0.050	0.071	0.057
	(3.140)				(0.129)	(0.535)	(0.777)	(0.702)
High Disp_adj	**1.284**	94	1.284	0.338	0.088	**0.118**	**0.167**	**0.120**
	(3.315)				(1.320)	**(1.871)**	**(2.749)**	**(2.577)**
H-L Disp_adj	0.038		0.038	0.037	0.075	0.069	0.096	0.063
	(0.366)				(0.742)	(0.680)	(0.954)	(0.696)

Note: This table displays the results of portfolio sorts based on ESG rating disagreement. We sort stocks in the sample into five groups based on the firm's ESG rating disagreement adjusted by the average rating disagreement in the firm's Fama French 12 industry. We report rating disagreement for the Total, Environmental, Social and Governance ratings. The panels display the portfolios with the smallest disagreement (Low *Disp_adj*), the highest disagreement (High *Disp_adj*), as well as the long-short portfolio between these portfolios (H-L *Disp_adj*). We report mean returns (Ret), the median number of observations for the high and low disagreement portfolios (N), the standard deviation of returns (Stddev), and the sharpe ratio (SR). In addition, we report CAPM, Fama-French 3 factor, Carhart 4 factor and Fama-French 5 factor alphas (columns 5-8). The sample includes 96 monthly time-series observations from January 2010 to December 2017. Portfolios are formed each January with disagreement values from December of the preceding year. Returns are reported in percent. *T*-statistics are reported in parentheses. *T*-statistics for the factor model alphas are based on Newey-West standard errors with 12 lags.

Table 6: Portfolio sorts on industry-adjusted ESG rating disagreement (adjusted for trading costs)

	(1) Ret	(2) N	(3) Stddev	(4) SR	(5) CAPM	(6) FF3	(7) Car	(8) FF5
					Panel A: Total			
Low Disp_adj	**1.119**	93	1.119	0.295	-0.089	-0.068	-0.056	-0.078
	(2.887)				(-1.288)	(-0.988)	(-0.820)	(-1.110)
High Disp_adj	**1.330**	94	1.330	0.354	0.139	**0.159**	**0.206**	0.162
	(3.465)				(1.500)	**(1.715)**	**(2.365)**	(1.567)
H-L Disp_adj	**0.181**		0.181	0.194	**0.196**	**0.195**	**0.227**	**0.203**
	(1.897)				**(1.890)**	**(1.906)**	**(2.370)**	**(1.830)**
					Panel B: Environmental			
Low Disp_adj	**1.178**	93	1.178	0.291	-0.097	-0.050	-0.023	-0.041
	(2.848)				(-0.977)	(-0.575)	(-0.276)	(-0.510)
High Disp_adj	**1.379**	93	1.379	0.355	**0.150**	**0.169**	**0.197**	**0.162**
	(3.474)				**(1.775)**	**(1.964)**	**(2.247)**	**(1.882)**
H-L Disp_adj	0.168		0.168	0.164	**0.214**	**0.185**	**0.185**	0.166
	(1.609)				**(2.155)**	**(1.861)**	**(1.939)**	(1.594)
					Panel C: Social			
Low Disp_adj	**1.270**	93	1.270	0.331	0.059	0.109	**0.126**	**0.102**
	(3.243)				(0.742)	(1.589)	**(1.758)**	**(1.690)**
High Disp_adj	**1.320**	93.5	1.320	0.333	0.065	0.089	**0.125**	0.075
	(3.260)				(0.681)	(1.052)	**(1.693)**	(0.985)
H-L Disp_adj	0.017		0.017	0.018	-0.029	-0.055	-0.038	-0.066
	(0.172)				(-0.443)	(-0.777)	(-0.630)	(-0.954)
					Panel D: Governance			
Low Disp_adj	**1.225**	93	1.225	0.312	-0.021	0.020	0.040	0.025
	(3.054)				(-0.201)	(0.203)	(0.421)	(0.297)
High Disp_adj	**1.266**	93	1.266	0.331	0.062	0.091	**0.141**	**0.090**
	(3.247)				(0.934)	(1.444)	**(2.274)**	**(1.916)**
H-L Disp_adj	0.005		0.005	0.005	0.045	0.033	0.060	0.022
	(0.044)				(0.401)	(0.302)	(0.562)	(0.227)

Note: This table displays the results of portfolio sorts based on ESG rating disagreement. We sort all stocks in the sample into five groups based on the firm's ESG rating disagreement adjusted by the average rating disagreement in the firm's Fama French 12 industry. We report rating disagreement for the Total, Environmental, Social and Governance ratings. The panels display the portfolios with the smallest disagreement (Low *Disp_adj*), the highest disagreement (High *Disp_adj*), as well as the long-short portfolio between these portfolios (H-L *Disp_adj*). We report mean returns (Ret), the median number of observations for the high and low disagreement portfolios (N), the standard deviation of returns (Stddev), and the sharpe ratio (SR). In addition, we report CAPM, Fama-French 3 factor, Carhart 4 factor and Fama-French 5 factor alphas (columns 5-8). The sample includes 96 monthly time-series observations from January 2010 to December 2017. Portfolios are formed each January with disagreement values from December of the preceding year. Returns are reported in percent. *T*-statistics are reported in parentheses. T-statistics for the factor model alphas are based on Newey-West standard errors with 12 lags. We adjust returns for trading costs according to Novy-Marx and Velikov (2016).

Table 7: Risk and ESG rating disagreement

	(1)	(2)	(3)	(4)
Pillars:	Total	Environmental	Social	Governance
Dependent Variable:	Total Volatility			
Disp	0.002	-0.001	**0.003**	-0.001
	(1.614)	(-0.667)	**(2.124)**	(-0.949)
Controls	Yes	Yes	Yes	Yes
N	40,519	40,365	40,519	40,519
Adjusted R^2	0.573	0.571	0.573	0.573
Dependent Variable:	Idiosyncratic Volatility			
Disp	**0.002**	-0.001	**0.003**	-0.001
	(1.710)	(-0.463)	**(2.223)**	(-0.819)
Controls	Yes	Yes	Yes	Yes
N	40,515	40,362	40,515	40,515
Adjusted R^2	0.477	0.474	0.477	0.476
Dependent Variable:	Beta			
Disp	0.075	-0.159	0.075	-0.088
	(0.690)	(-1.469)	(0.777)	(-0.959)
Controls	Yes	Yes	Yes	Yes
N	40,515	40,362	40,515	40,515
Adjusted R^2	0.441	0.442	0.441	0.441
Dependent Variable:	LPM			
Disp	**0.002**	0.000	**0.002**	-0.001
	(1.915)	(-0.271)	**(1.989)**	(-0.865)
Controls	Yes	Yes	Yes	Yes
N	40,515	40,362	40,515	40,515
Adjusted R^2	0.519	0.517	0.519	0.518

Note: This table displays the results of pooled panel regressions of monthly stock level risk measures on ESG rating disagreement. The results are separated into four panels: We use four different risk measures as dependent variables: total volatility log transformed *(Total Volatility)*, idiosyncratic volatility log transformed *(Idiosyncratic Volatility)*, stock market beta (*Beta*), and the lower partial moment log transformed (*LPM*). For the calculation of idiosyncratic volatility and stock market beta, we employ the CAPM. We measure ESG rating disagreement by the standard deviation of ratings available for a given firm at a given point in time (*Disp*). We also include industry-month fixed effects and control for standard characteristics (not reported) that have been found to explain volatility (see, for example, Dennis and Stickland 2004), namely market capitalization (Banz 1981), leverage, business segment (dummy variable, which is one for multi-segment firms), percentage of institutional ownership, the ratio of mutual fund ownership to total institutional ownership, and turnover. *T*-statistics based on double clustered standard errors (month and firm) are reported in parentheses.

References

Anderson, E. W., Ghysels, E. and Juergens, J. L.: 2005, Do heterogeneous beliefs matter for asset pricing?, *The Review of Financial Studies* 18(3), 875–924.

Anderson, E. W., Ghysels, E. and Juergens, J. L.: 2009, The impact of risk and uncertainty on expected returns. *Journal of Financial Economics* 94(2), 233–263.

Ang, A., Hodrick, R. J., Xing, Y. and Zhang, X.: 2006, The cross-section of volatility and expected returns, *The Journal of Finance* 61(1), 259–299.

Atmaz, A. and Basak, S.: 2018, Belief dispersion in the stock market, *The Journal of Finance* 73(3), 1225–1279.

Avramov, D., Cheng, S., Lioui, A. and Tarelli, A.: 2021, Sustainable investing with ESG rating uncertainty. Working Paper.

Avramov, D., Chordia, T., Jostova, G. and Philipov, A.: 2009, Dispersion in analysts' earnings forecasts and credit rating, *Journal of Financial Economics* 91(1), 83–101.

Banz, R. W.: 1981, The relationship between return and market value of common stocks, *Journal of Financial Economics* 9(1), 3–18.

Berg, F., Koelbel, J. F. and Rigobon, R.: 2020, Aggregate Confusion: The Divergence of ESG Ratings. Working Paper.

Bolton, P., Kacperczyk, A.: 2021, Do Investors care about carbon risk?, Forthcoming. *Journal of Financial Economics*.

Bouten, L., Cho, C. H., Michelon, G. and Roberts, R. W.: 2017, CSR performance proxies in large-sample studies: 'Umbrella advocates', construct clarity and the 'validity police'. Working Paper.

Carhart, M. M. (1997). On persistence in mutual fund performance. The Journal of finance, 52(1), 57-82.

Chatterji, A. K., Durand, R., Levine, D. I. and Touboul, S.: 2016, Do ratings of firms converge? Implications for managers, investors and strategy researchers, *Strategic Management Journal* 37(8), 1597–1614.

Chatterji, A. K., Levine, D. I. and Toffel, M. W.: 2009, How well do social ratings actually measure corporate social responsibility?, *Journal of Economics & Management Strategy* 18(1), 125–169. 00262.

Chordia, T., Subrahmanyam, A., and Tong, Q.: 2014, Have capital market anomalies attenuated in the recent era of high liquidity and trading activity?. Journal of Accounting and Economics, 58(1), 41-58.

Christensen, D., Serafeim, G. and Sikochi, A.: 2021, Why is Corporate Virtue in the Eye of the Beholder? The Case of ESG Ratings. Forthcoming at *Journal of Accounting Research*.

Delmas, M. A., Etzion, D. and Nairn-Birch, N.: 2013, Triangulating environmental performance: What do corporate social responsibility ratings really capture?, *Academy of Management Perspectives* 27(3), 255–267.

Dennis, P. J. and Stickland, D.: 2004, The determinants of idiosyncratic volatility. Working Paper.

Diether, K. B., Malloy, C. J. and Scherbina, A.: 2002, Differences of opinion and the cross section of stock returns, *The Journal of Finance* 57(5), 2113–2141.

Doyle, T. M.: 2018, Ratings that don't rate – the subjective view of ESG rating agencies, *Technical report*, ACCF American Council for Capital Formation. URL: https://bit.ly/2LBwvky

Dyck, A., Lins, K. V., Roth, L. and Wagner, H. F.: 2019, Do institutional investors drive corporate social responsibility? International evidence, *Journal of Financial Economics* 131(3), 693–714.

Dunn, J., Fitzgibbons, S., and Pomorski, L., 2018. Assessing risk through environmental, social and governance exposures. Journal of Investment Management, 16(1), 4-17.

Eccles, R. G., Lee, L.-E. and Stroehle, J. C.: 2019, The social origins of ESG? An analysis of Innovest and KLD. Working Paper.

Fama, E. F. and French, K. R.: 1995, Size and book-to-market factors in earnings and returns, *The Journal of Finance* 50(1), 131–155.

Fama, E. F., & French, K. R.: 2015. A five-factor asset pricing model, *Journal of Financial Economics* 116(1), 1–22.

Fama, E. F. and MacBeth, J. D.: 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81(3), 607–636.

Frazzini, A. and Pedersen, L. H.: 2014, Betting against beta, *Journal of Financial Economics* 111(1), 1–25.

Friede, G., Busch, T., and Bassen, A.: 2015, ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, 5(4), 210-233.

Gibson, R., Krueger, P., and Mitali, S.: 2020, The sustainability footprint of institutional investors, *Swiss Finance Institute Research Paper* (17-05).

Gibson, R., Glossner, S., Krueger, P., Matos, P. and Steffen, T.: 2021, Do Responsible Investors Invest Responsibly? University of Geneva and University of Virginia.

Green, J., Hand, J.R. and Zhang, X.F., 2017. The characteristics that provide independent information about average US monthly stock returns. *The Review of Financial Studies*, 30(12), pp.4389-4436.

GSIA Global Sustainable Investment Alliance: 2016, Global sustainable investment review 2016, https://bit.ly/2U1OuoF.

Hasbrouck, J.: 2009, Trading costs and returns for U.S. equities: Estimating effective costs from daily data, *The Journal of Finance* 64(3), 1445–1477.

Hoepner, A. G., Oikonomou, I., Sautner, Z., Starks, L. T. and Zhou, X.: 2019, ESG shareholder engagement and downside risk. Working Paper.

Hong, H. and Kostovetsky, L.: 2012, Red and blue investing: Values and finance, *Journal of Financial Economics* 103(1), 1–19.

Jacobs, H. and Müller, S.: 2020, Anomalies across the globe: Once public, no longer existent?. *Journal of Financial Economics* 135(1), 213–230.

Jegadeesh, N. and Titman, S.: 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *The Journal of Finance* 48(1), 65–91.

Kogan, L and Wang, T.: 2003, A simple theory of asset pricing under model uncertainty. MIT Working Paper.

Krüger, P.: 2015, Corporate goodness and shareholder wealth, *Journal of Financial Economics* 115(2), 304–329.

Liang, H. and Renneboog, L.: 2017, On the foundations of corporate social responsibility, *The Journal of Finance* 72(2), 853–910.

Lins, K. V., Servaes, H. and Tamayo, A.: 2017, Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis, *The Journal of Finance* 72(4), 1785–1824.

Lopez-de-Silanes, F., McCahery, J. A. and Pudschedl, P. C.: 2019, ESG performance and disclosure: A cross-country analysis. ECGI Working Paper.

Matos, P.: 2020, ESG and Responsible Institutional Investing Around the World: A Critical Review. CFA Institute Research Foundation Literature Reviews, May 2020.

Mackintosh, J.: 2018, Is Tesla or Exxon more sustainable? It depends whom you ask, *Wall Street Journal*. URL: *https://on.wsj.com/2MQCC4m*

McLean, R.D. and Pontiff, J.: 2016, Does academic research destroy stock return predictability?. *The Journal of Finance*, 71(1), pp.5-32.

Novy-Marx, R.: 2013, The other side of value: The gross profitability premium, *Journal of Financial Economics* 108(1), 1–28.

Novy-Marx, R., and Velikov, M.: 2016, A taxonomy of anomalies and their trading costs, *The Review of Financial Studies* 29(1), 104–147.

PRI Principles for Responsible Investment: 2018, Annual report 2018, https://bit.ly/2NsU8xv.

Roll, R.: 1984, A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance*, 39(4), 1127-1139.

USSIF Forum for Sustainable and Responsible Investment: 2020, Report on U.S. Sustainable and Impact Investing Trends 2020, Washington DC.

Viale A.M., L. Garcia-Feijoo and Giannetti, A.: 2014, Safety first, learning under ambiguity, and the cross-section of returns, *Review of Asset Pricing Studies* 4(1), 118–159.

Wigglesworth, R.: 2018, Rating agencies using green criteria suffer from 'inherent biases', *Financial Times*. URL: https://on.ft.com/2H5E0kj

Wong, C., A. Brackley, and E. Petroy.: 2019, Rate the raters 2019: Expert views on ESG ratings.

Appendix A: Sample selection, financial data, dataset matching and variable definitions

A.1 Sample selection

To test our hypotheses, we construct a representative and homogeneous sample over the longest possible time period. We face the challenge that the availability of ESG data is restricted in both the cross-section and the time-series. In other words, ESG data is often only available for the largest firms and for more recent years. To use a sample as homogeneous as possible and to maximize the number of available ESG ratings per firm, we restrict ourselves to firms belonging to the S&P 500 and consider a sample period of eight years going from 2010 to 2017.
See Table A.1 for an overview of the variables used in this study.

[Appendix Table A.1 about here.]

A.2 Financial data

We use financial data from the Center for Research in Security Prices (CRSP) and accounting data from Standard & Poor's Compustat. For each stock, we calculate idiosyncratic volatility, total volatility, and the stock market beta at the end of each month using up to 250 daily observations (we require a minimum of 60 daily observations). We calculate market capitalization as (adjusted) total shares outstanding times stock price, both at the end of the month. The momentum signal at time t is calculated as the continuously compounded returns from month $t - 2$ to month $t - 12$. Book value of equity is the sum of shareholders' equity, deferred taxes and investment tax credit minus preferred stock.[16] Only firms with a positive

[16] If available, we use the redemption value as preferred stock. Otherwise, we use the liquidating value or, if the liquidation value is also not available, the carrying value.

book value are selected into the sample. Following Novy-Marx (2013), gross profitability is calculated as total revenues (revt) minus cost of goods sold (cogs), divided by total assets (at). In addition, we also match the dispersion in EPS forecasts for one year ahead earnings from IBES (Diether et al. 2002).

A.3 Dataset matching

A big challenge for constructing a dataset from many sub-datasets is to properly match the different datasets. We match on three identifiers: (1) CUSIP, (2) ISIN, and (3) company name. The CUSIP code is available for all providers, except Inrate.[17] However, since the ISIN code is available for Inrate, we extract the CUSIP code from the ISIN code. Note also, that we only use the first six CUSIP characters for matching (known as the *issuer* identifier). The characters seven and eight identify the specific issue (for example *10* indicates common equity), and the ninth character is a check digit. The ISIN code is available for all providers except MSCI KLD. For the CRSP/Compustat data we retrieve the ISIN number from the CUSIP code and the current ISO country code of incorporation (fic).[18] To do the merge with the company names, we first convert the original names of the providers, by using some commonly used abbreviations to avoid rather trivial mismatches. We use the unique union of all three matching procedures to compile our sample.

[17] For the MSCI KLD dataset there seems to be some issues with the CUSIP code. The codes do not always have the same number of characters, and it seems that leading zeros are often truncated. Therefore, we re-fill leading zeros if the number of characters is less than eight. Then we add the self-computed check digit to the code if the eighth number is not the would-be check digit if there would be an additional leading zero (in that case we add a leading zero) or the last two characters consist of commonly used issue codes.

[18] For US stocks the ISIN number is composed of the country code (first two characters), the CUSIP code (characters three to eleven), and a check digit.

To construct the sample, we also require that at least three rating observations are available for each company. This choice provides us with an internally consistent sample, and, in addition, it is not overly restrictive.

In addition, we use a monthly frequency for our sample. Asset 4, Sustainalytics, FTSE and MSCI IVA already provide data at a monthly frequency; Inrate provides ratings update on a semi-annual basis for the years 2015 and 2016; and Bloomberg and MSCI KLD provide data on a yearly frequency. To convert from a semi-annual or annual frequency, we simply use the respective annual or semi-annual value for the whole time period. Note that most ratings (also for the providers with a monthly frequency) change rather infrequently, with most ratings being constant for about one year, but also for longer periods.[19]

[19] Since the providers change their ratings at different points in time, we argue that for our purposes it makes sense to use a monthly frequency.

Appendix Table A.1: Variables Overview

Variables	Description	Details	Source
ESG rating disagreement variables			
Disp	Standard deviation of all firm-level ratings	To compute standard deviations, we adjust the raw ratings as follows: we calculate the percentile ranks and use these as adjusted scores.	Thomson Reuters/Refinitiv, Sustainalytics, Inrate, Bloomberg, FTSE MSCI
Disp_adj	Industry-adjusted standard deviation of firm-level ratings disagreement	A firm's ESG rating disagreement adjusted by the average rating disagreement in the firm's Fama French 12 industry.	Thomson Reuters/Refinitiv, Sustainalytics, Inrate, Bloomberg, FTSE MSCI, Fama and French Data Library
Additional independent variable(s)			
Return	Stock returns	Monthly stock returns	CRSP
Control variables			
TAN	Tangibility	Property, plant, and equipment (PPENT) divided by total Assets (AT).	Compustat
CR	Current Ratio	Current assets (ACT) divided by current liabilities (LCT).	Compustat
LEV	Leverage	Long term debt (DLTT) plus debt in current liabilities (DLC) divided by total assets (AT).	Compustat
GP	Gross Profitability	Revenues (REVT) minus costs of goods sold (COGS) divided by Total Assets (AT).	Compustat
HHI	Herfindahl-Hirschman Index (HHI) based on book equity	The HHI measures industry concentration, by using book equity and the 2-digit SIC level.	Compustat
MSEG	Multi-Segment	Dummy variable, which is one if the firm operates in more than one segment.	Compustat Segments Data
NCR	Missing Credit Rating	Dummy variable, which is one if there is no credit rating available.	Compustat Company S&P Credit Ratings
IO	Institutional Ownership	Percentage of institutional ownership.	Thomson Reuters Institutional (13f) Holdings
NoA	Number of Analysts	Number of analysts, based on IBES summary files.	IBES
StdEPS	Dispersion of analyst forecasts of the firm's one year ahead earnings forecasts	Dispersion of analyst forecasts of the firm's one year ahead earnings forecasts, measured by standard deviation.	IBES
BM	Book-to-Market Ratio	Book equity (shareholders' equity (SEQ) plus deferred taxes (TXDB) plus investment tax credit (ITCB) minus preferred stock (which is either redemption value (PSTKRV), liquidation value (PSTKL) or carrying value (PSTK), based on availability)) divided by Market Capitalization.	Compustat, CRSP
ME	Market Capitalization	Absolute value of stock price (PRC) times shares outstanding (SHROUT).	CRSP
Momentum	Momentum	Cumulative returns of the most recent 12 month, excluding the most recent one for each firm (from month t-12 to t-2).	CRSP
TVol	Total Volatility	Standard Deviation computed from the most recent 250 daily return observations.	CRSP
Beta	Firm's beta	Market beta computed from the most recent 250 daily return observations.	CRSP

Continued on next page

Downside risk measures			
LPM	Lower partial moment (log-transformed)	The lower partial moment is the square root of the standard deviation of the negative return part of the distribution. For details, see Hoepner et al. (2019).	CRSP

Note: This table provides an overview of the variables used in this study. We classify the variables into four groups: ESG rating disagreement variables, additional independent variable(s), control variables and downside risk measures.

Appendix Table A.2: Summary statistics

Variable	N	Mean	StdDev	Min	Max	Median	Skew	Kurt
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ESG rating disagreement variables							
Disp (T)	51,206	0.200	0.085	0.000	0.560	0.194	0.386	-0.055
Disp (E)	50,904	0.204	0.081	0.000	0.547	0.202	0.287	0.075
Disp (S)	50,877	0.222	0.083	0.000	0.544	0.220	0.131	-0.285
Disp (G)	50,877	0.241	0.081	0.000	0.535	0.242	-0.007	-0.315
	Other variables							
Return	51,152	0.012	0.080	-0.748	1.274	0.013	0.388	6.796
ME	51,178	0.030	0.052	0.000	0.882	0.013	4.992	38.167
BM	49,783	0.762	2.381	0.001	108.763	0.444	20.718	621.428
GP	49,783	0.287	0.219	-1.143	1.405	0.253	1.035	2.459
Momentum	47,736	0.150	0.305	-0.925	4.870	0.139	1.320	10.567
StdEPS	44,310	0.822	17.904	0.000	961.040	0.060	31.516	1147.771
Beta	51,094	1.075	0.394	-1.738	3.787	1.039	0.712	1.461
TVol	51,178	0.018	0.009	0.001	0.226	0.016	4.899	65.671
TAN	49,304	0.255	0.244	0.000	0.946	0.160	1.022	-0.102
CR	42,716	1.832	1.188	0.205	11.966	1.490	2.673	11.596
LEV	49,585	0.246	0.157	0.000	0.960	0.238	0.591	0.561
HHI	49,783	0.074	0.078	0.009	0.744	0.044	2.810	11.632
MSEG	18,212	0.562	0.496	0.000	1.000	1.000	-0.249	-1.938
NCR	51,206	0.199	0.399	0.000	1.000	0.000	1.511	0.284
IO	50,811	0.751	0.221	0.000	1.748	0.801	-1.811	3.800
NoA	48,920	0.189	0.078	0.010	0.560	0.180	0.445	0.481
IVol	51,094	0.014	0.008	0.001	0.226	0.012	6.994	135.876
LPM	51,094	0.012	0.006	0.001	0.103	0.011	3.054	21.447

Note: This table shows descriptive statistics of all variables used in the study. We report the number of observations (N), the sample means (Mean) and standard deviations (StdDev) as well as minimum (Min), maximum (Max), medians (Median), skewness (Skew) and Kurtosis (Kurt). We show the descriptive statistics for the following variables: Ratings disagreement for each of the four pillars (Total, Environmental, Social, and Governance pillar): Disp (T/E/S/G). A detailed description of the variables is given in Table A.1.

Appendix B: Robustness check – sample without FTSE and Inrate

Appendix Table B.1: Stock returns and ESG rating disagreement – sample without FTSE and Inrate

Dependent Variable:	Returns			
	(1)	(2)	(3)	(4)
Pillars:	Total	Environmental	Social	Governance
Disp	**0.626**	**1.217**	**0.698**	-0.376
	(1.762)	**(3.005)**	**(1.720)**	(-1.023)
ME	-0.131	-0.108	-0.129	-0.111
	(-0.710)	(-0.579)	(-0.694)	(-0.592)
BM	0.218	0.240	0.228	0.229
	(0.983)	(1.076)	(1.026)	(1.036)
GP	0.225	0.256	0.240	0.219
	(0.751)	(0.860)	(0.799)	(0.735)
Momentum	0.361	0.348	0.357	0.361
	(1.114)	(1.063)	(1.102)	(1.120)
StdEPS	-0.230	-0.194	-0.234	-0.225
	(-1.325)	(-1.119)	(-1.355)	(-1.298)
Beta	0.158	0.202	0.158	0.155
	(0.316)	(0.403)	(0.317)	(0.310)
TVol	-0.078	-0.101	-0.082	-0.073
	(-0.220)	(-0.288)	(-0.231)	(-0.206)
Industry * Month FE	Yes	Yes	Yes	Yes
N	42,032	41,738	42,032	42,032
Adjusted R^2	0.347	0.348	0.347	0.347

Note: This table displays the results of pooled panel regressions of monthly stock returns on ESG rating disagreement. The first row reports the results for disagreement in the total ESG rating and the rows (2) – (4) report the results for the E, S, and G pillar separately. The dependent variable *Returns* is the firm's monthly stock return. We measure ESG rating disagreement by the standard deviation of ratings available for a given firm at a given point in time (*Disp*). We also include industry-month fixed effects and control for standard characteristics that have been found to explain stock returns, namely market capitalization (Banz 1981), book-to-market (Fama and French 1995), gross profitability (Novy-Marx 2013), momentum (Jegadeesh and Titman 1993), the dispersion of analyst forecasts of the firm's one year ahead earnings forecasts (Diether et al. 2002), the firm's beta (Frazzini and Pedersen 2014), and total volatility (Ang et al. 2006). *t*-statistics based on double clustered standard errors (month and firm) are reported in parentheses.

Swiss Finance Institute

Swiss Finance Institute (SFI) is the national center for fundamental research, doctoral training, knowledge exchange, and continuing education in the fields of banking and finance. SFI's mission is to grow knowledge capital for the Swiss financial marketplace. Created in 2006 as a public–private partnership, SFI is a common initiative of the Swiss finance industry, leading Swiss universities, and the Swiss Confederation.

swiss:finance:institute

c/o University of Geneva, Bd. Du Pont d'Arve 42, CH-1211 Geneva 4
T +41 22 379 84 71, rps@sfi.ch, www.sfi.ch

european corporate governance institute

about ECGI

The European Corporate Governance Institute has been established to improve *corporate governance through fostering independent scientific research and related activities.*

The ECGI will produce and disseminate high quality research while remaining close to the concerns and interests of corporate, financial and public policy makers. It will draw on the expertise of scholars from numerous countries and bring together a critical mass of expertise and interest to bear on this important subject.

The views expressed in this working paper are those of the authors, not those of the ECGI or its members.

www.ecgi.global

european corporate governance institute

ECGI Working Paper Series in Finance

Editorial Board

Editor	Mike Burkart, Professor of Finance, London School of Economics and Political Science
Consulting Editors	Franklin Allen, Nippon Life Professor of Finance, Professor of Economics, The Wharton School of the University of Pennsylvania
	Julian Franks, Professor of Finance, London Business School
	Marco Pagano, Professor of Economics, Facoltà di Economia Università di Napoli Federico II
	Xavier Vives, Professor of Economics and Financial Management, IESE Business School, University of Navarra
	Luigi Zingales, Robert C. McCormack Professor of Entrepreneurship and Finance, University of Chicago, Booth School of Business
Editorial Assistant	Úna Daly, Working Paper Series Manager

www.ecgi.global/content/working-papers

european corporate governance institute

Electronic Access to the Working Paper Series

The full set of ECGI working papers can be accessed through the Institute's Web-site (www.ecgi.global/content/working-papers) or SSRN:

Finance Paper Series	http://www.ssrn.com/link/ECGI-Fin.html
Law Paper Series	http://www.ssrn.com/link/ECGI-Law.html

www.ecgi.global/content/working-papers